ASISTENTES DE APRENDIZAJE BASADOS EN INTELIGENCIA ARTIFICIAL: PRINCIPIOS DE SEGURIDAD Y EXPERIENCIAS DE IMPLEMENTACIÓN EN EDUCACIÓN SUPERIOR

Maria Jose Casañ
UPC Universidad Politécnica de Cataluña - Barcelona Tech, Barcelona, España

Marc Alier
UPC Universidad Politécnica de Cataluña - Barcelona Tech, Barcelona, España

Juanan Pereira
Facultad de Informática, UPV/EHU, San Sebastián, España

Francisco José García-Peñalvo
Instituto Universitario de Ciencias de la Educación (IUCE)
Universidad de Salamanca, Salamanca, España

1. INTRODUCCIÓN

Desde finales de 2022, la adopción de herramientas de Inteligencia Artificial (IA) basadas en aprendizaje automático en el ámbito educativo ha experimentado un súbito crecimiento. La apertura al gran público de herramientas de IA Generativa (IAGen) (García-Peñalvo & Vázquez-Ingelmo, 2023; Jovanović & Campbell, 2022) como ChatGPT o Dall-e han demostrado como artefactos experimentales de investigación en IA pueden convertirse rápidamente en aplicaciones prácticas dentro de los entornos educativos (Alier-Forment & Llorens-Largo, 2023). Esta transformación ha despertado un interés creciente entre docentes y administradores de centros educativos, quienes han comenzado a explorar de manera acelerada las posibles implicaciones de la IA en educación, así como las respuestas estratégicas necesarias para aprovechar estas tecnologías de manera efectiva (García-Peñalvo, 2024).

Un claro ejemplo de esta tendencia es ChatGPT, chatbot basado en el Large Language Model (LLM) GPT-3 (Brown et al., 2020) que OpenAI a finales de 2022 abrió al público como demo de un producto de investigación. De la noche a la mañana, ChatGPT obtuvo millones de usuarios y puso a la IAGen en el centro de atención del público general,

convirtiéndose en un referente. Estas herramientas, entre las que se incluyen variantes como Claude 3 (Anthropic, 2024) de Anthropic y Gemini (Pichai & Hassabis, 2024) de Google, han proliferado desde entonces. A partir de 2023 el campo en ebullición de la IAGen empieza a ofrecer una gama creciente de funcionalidades, tanto a los usuarios finales como a desarrolladores a través de interfaces de programación y plataformas de código abierto, como es el caso de Llama 3 (Meta, 2024) de Meta y Mixtral 8x22B (Mistral AI team, 2024a) o Pixtral Large (Mistral AI team, 2024b) de Mistral. Estos avances han generado un debate en torno a las métricas de rendimiento y las posibles aplicaciones de estas tecnologías en contextos educativos.

El auge de ChatGPT no solo ha impactado las prácticas docentes tradicionales, sino que también ha impulsado a los educadores a investigar cómo integrar estas herramientas en diversas tareas educativas, que van desde la automatización de procesos administrativos hasta la mejora de las experiencias de aprendizaje (García-Peñalvo, Llorens-Largo, et al., 2024). Las capacidades de estas herramientas abarcan una amplia gama de aplicaciones, como la creación de contenido, la facilitación de discusiones y la personalización de la realimentación. Además, su uso se extiende al apoyo en investigaciones académicas, sugiriendo metodologías, analizando datos y mejorando la redacción científica (Bahrini et al., 2023; Crawford et al., 2023; García-Peñalvo, 2023).

Este panorama de innovación tecnológica abre múltiples oportunidades para rediseñar los métodos de enseñanza y aprendizaje, introduciendo nuevas formas de interacción y potenciando la personalización del proceso educativo. La IAGen, con su capacidad de adaptarse a las necesidades de los estudiantes y docentes, promete revolucionar el modo en que se concibe y practica la educación en los próximos años. No obstante, la integración de la IA en la educación no puede hacerse de cualquier manera, porque existen aún numerosos problemas y retos que se han observado en experiencias de integración de IA en educación. Los más importantes son:

- Calidad de los prompts. La efectividad de los LLM depende en gran medida de la calidad de los prompts que les dan los usuarios (Morales-Chan, 2023). Crear prompts de alta calidad no es tarea sencilla y se asemeja más a un arte que a una disciplina técnica (Henley et al., 2024).
- Respuestas variables. Los LLM pueden generar respuestas de calidad variable (Yang et al., 2024), sobre todo en áreas donde los datos de entrenamiento son limitados o incompletos. Esto puede ser problemático en la educación, donde la precisión es crucial.
- Alucinaciones de la IA. Los LLM pueden producir lo que se conoce como "alucinaciones", es decir, contenido que parece creíble, pero que es falso o irrelevante (Huang et al., 2024). Aunque esto representa un problema, también se puede utilizar como una oportunidad educativa. Los docentes pueden fomentar el pensamiento crítico y la alfabetización mediática desafiando al estudiantado a: a) Detectar e identificar las alucinaciones en un texto, b) Explicar por qué el modelo generó información incorrecta, c)

Discutir las consecuencias de confiar en información inexacta, y d) Desarrollar estrategias para verificar la precisión de la información generada por IA.

- Privacidad, seguridad y aspectos legales. Usar IA puede poner en riesgo datos sensibles, ya que muchas empresas de IA generativa no garantizan que las conversaciones con sus chatbots no se usen para otros fines, como el entrenamiento de nuevos modelos, salvo en planes empresariales de pago (Iskender, 2023).
- Dependencia de la tecnología. Existe el riesgo de que el estudiantado se vuelva demasiado dependiente de estas herramientas, lo que podría disminuir su creatividad y capacidad de pensamiento crítico (Duong et al., 2024). Sin embargo, si se utilizan correctamente, estas herramientas pueden potenciar dichas habilidades (Vartiainen & Tedre, 2023).
- Sesgos ocultos. Las respuestas de los LLM pueden reflejar los sesgos presentes en los datos con los que fueron entrenados (Kamath et al., 2024), lo que puede generar respuestas parciales o injustas.
- Falta de interacción humana. Aunque los chatbots inteligentes pueden ayudar en el aprendizaje, no pueden reemplazar la interacción humana, que es esencial para el desarrollo del estudiantado (Choi et al., 2023).
- Cuestiones éticas. El uso de LLM plantea preocupaciones éticas como el plagio, la falta de autoría del contenido generado (Johinke et al., 2023) y el acceso desigual a estas herramientas, especialmente en sus versiones premium de pago (Cotton et al., 2024).
- Desajuste con los contextos educativos. Los LLM no están integrados en las dinámicas educativas estructuradas, como las actividades grupales o individuales, la supervisión docente o los análisis de aprendizaje (García-Peñalvo, 2024). Tampoco están ajustados al contenido específico de los cursos ni a los modelos pedagógicos. Esto puede llevar a problemas como la generación de información incorrecta o engañosa (Fonseca-Escudero et al., 2023).
- Un ejemplo claro de estas dificultades es el uso de Khanmigo, una IA de Khan Academy diseñada para ayudar en la educación, que no ha cumplido del todo con las expectativas, especialmente en tareas específicas como el aprendizaje de idiomas (Shetye, 2024).

2. DIRECTRICES PARA UTILIZAR IA DE FORMA SEGURA EN EDUCACIÓN

Para poder integrar la IA en la educación de forma segura y cumpliendo con las normativas de privacidad de la Unión Europea y asegurando que la IA está alineada con los valores, estrategia y prácticas de la institución educativa, los autores proponen una serie de directrices que aseguran que las aplicaciones de IA estén alineadas con las estrategias educativas y mantienen los niveles necesarios de seguridad, precisión e integridad desde el punto de vista ético. Siguiendo estos principios, las instituciones educativas pueden aprovechar el potencial de la IA mientras mitigan riesgos relacionados con la privacidad, el mal uso y la exactitud de la información, aspectos esenciales para salvaguardar la calidad y equidad del proceso de aprendizaje.

2.1. Principios de IA segura en educación

La atención y preocupación por el impacto real, potencial y, a veces, imaginado de la IAGen en todos los ámbitos de la sociedad, incluida la educación, han causado en la Unión Europea y otros países, como China, la creación de legislación específica para la IA. Pero para poder cumplir con las nuevas legislaciones, códigos deontológicos educativos y satisfacer a comités de ética, para cualquier tecnología es necesario disponer de principios accionables de seguridad en la aplicación de esta. Es por esto por lo que los autores proponen siete principios para la aplicación de la IAGen de forma segura en entornos educativos (García-Peñalvo, Alier, et al., 2024).

Estos principios se basan en una observación de las legislaciones relevantes, como las normativas de privacidad y regulaciones específicas para la IA, aprobadas por la Unión Europea (European Parliament, 2024; European Parliament & Council of the European Union, 2016), pero también prestando atención a los principios tecnológicos en los que se basan estas herramientas. Pues la IAGen consiste en una familia de tecnologías y la seguridad debe partir de una comprensión técnica. Alier, García-Peñalvo y Camba (2024) proponen una definición clara de lo que significa una "IA Segura en la Educación" y presentan cinco principios (que en García-Peñalvo, Alier, Pereira y Casañ (2024) se amplían a siete) que deben cumplir las aplicaciones de IA para su uso en entornos educativos. Estos principios son aplicables y permiten una evaluación técnica de las estrategias de integración de IAGen en entornos educativos.

Es importante destacar que se hace referencia a "Sistema de IA" (generativa) y no a una "caja negra" que responde a peticiones como un oráculo. Toda aplicación de IAGen va a tener componentes muy avanzados basados en redes neuronales y modelos de aprendizaje automático, y otras partes de informática más clásica: servidores, redes de comunicación, y componentes de software tradicional (métodos de autenticación, protocolos, bases de datos, aplicaciones web, etc.).

Los principios propuestos son los siguientes:

-(SAIE1)Garantiza la confidencialidad. El sistema de IA debe asegurar la protección y confidencialidad de todos los datos de los estudiantes, incluyendo identidades, roles, expedientes académicos e interacciones.

-(SAIE2)Está alineada con las estrategias educativas. Las herramientas de IA deben estar en sintonía con las estrategias institucionales y las políticas de gobernanza tecnológica para apoyar los objetivos educativos y cumplir con los estándares operativos. Por ejemplo, deben facilitar el aprendizaje y la creación de contenidos, pero a la vez estar diseñadas para evitar el uso indebido, como hacer trampas o eludir medidas de integridad académica. El sistema no debe ofrecer soluciones a tareas o facilitar la paráfrasis para esquivar los controles de plagio.

-(SAIE3)Se ajusta a las prácticas didácticas. Las aplicaciones de IA deben seguir parámetros educativos predeterminados cuando se despliegan en entornos educativos. Imagínese una aplicación de IA que se usa en una clase de matemáticas, concretamente para ayudar en la resolución de ecuaciones. La IA, en lugar de simplemente resolver cualquier problema que le pidan, sigue las directrices establecidas por el profesorado. Por ejemplo, no da directamente la solución de la ecuación, sino que sigue un método, ofreciendo pistas o recordando los pasos que se han explicado en clase.

-(SAIE4)Precisión y minimización de errores. Aunque los modelos se entrenan con grandes repositorios de datos, existe el riesgo de que proporcionen información incorrecta – debido a errores o sesgos en los datos de entrenamiento – o alucinaciones. Un sistema de IA seguro debe priorizar la precisión y relevancia de sus respuestas, lo que es más factible en contextos de aplicación claramente definidos.

-(SAIE5)Interfaz comprensible y comportamiento adecuado. El sistema de IA debe presentarse de manera comprensible para estudiantes y profesores, clarificando sus usos previstos y limitaciones. Imagínese una herramienta de IA que ayuda a un grupo de estudiantes a buscar información sobre un periodo histórico. La herramienta, antes de cada búsqueda, muestra un mensaje breve que explica su propósito, por ejemplo, "Esta IA te ayudará a encontrar información histórica relevante y confiable, pero revisa siempre las fuentes y consulta con tu profesor/a en caso de duda". Esto ayuda al estudiantado a entender que, aunque la IA puede dar respuestas útiles, no es infalible y tiene ciertas limitaciones.

-(SAIE6)Supervisión humana y responsabilidad. Las herramientas de IA en la educación deben complementar, no reemplazar, a los educadores humanos. Si bien la IA puede asistir en tareas administrativas como la corrección o la retroalimentación, las decisiones deben estar siempre bajo la supervisión de personas. Las decisiones generadas por IA deben ser explicables. El estudiantado debe tener derecho a apelar dichas decisiones mediante procesos supervisados por humanos. Esto garantiza la equidad, mantiene el papel de mentoría del profesorado y protege la integridad del proceso educativo.

-(SAIE7)Entrenamiento ético y transparencia. Los modelos de IA utilizados en la educación deben entrenarse de manera ética, con un compromiso claro con la transparencia respecto a las fuentes de datos de entrenamiento y las metodologías empleadas. Es fundamental que estos modelos minimicen los sesgos y ofrezcan transparencia sobre sus procesos de formación, permitiendo a educadores y estudiantes entender las limitaciones de los resultados generados por la IA.

2.2. Implicaciones de los principios de IA segura en educación

➢El principio SAIE1 (garantía de confidencialidad) exige que la institución educativa tenga control sobre la herramienta de IA para garantizar la privacidad y confidencialidad de los estudiantes. Esto puede lograrse operando toda la tecnología de manera interna o asegurando la privacidad en los acuerdos con los proveedores de IA. Por tanto, el uso de herramientas gratuitas que obligan a los estudiantes a registrarse, como https://chatgpt.com, no debería ser obligatorio. El estudiantado puede hacerlo por decisión propia, pero no debería verse obligado a ello para completar tareas educativas. La revisión de la literatura muestra que la investigación primaria rara vez aborda los problemas de privacidad, como la protección de datos durante la recopilación en entornos educativos, por lo que existe la necesidad de mejorar los marcos éticos (Alam & Mohanty, 2022; Fichten et al., 2021).

➢El principio SAIE2 (alineación con las estrategias educativas) plantea tensiones con el uso de herramientas de propósito general como ChatGPT, diseñadas para múltiples casos de uso. Estas herramientas de propósito general pueden no encajar bien a nivel institucional por varias razones:

-La complejidad de usar un chatbot basado en un LLM es engañosa. Diseñar buenos prompts se está revelando como una tarea muy compleja (Willison, 2023). Agregar complejidad al proceso de aprendizaje no es una buena práctica pedagógica, ya que aumenta la carga cognitiva del estudiantado (Chen et al., 2023).

-Los chatbots basados en LLM están afinados para seguir instrucciones del usuario, por lo que evitar su uso para trampas o plagio es casi imposible (González-Geraldo & Ortega-López, 2024).

-Los LLM siempre proporcionan una respuesta, pero la calidad de estas respuestas varía mucho. El estudiantado, que puede no tener suficiente experiencia, podría verse engañado por respuestas incorrectas o irrelevantes (alucinaciones) (Perković et al., 2024).

➢El principio SAIE3 (alineación con las prácticas didácticas) introduce los mismos problemas que SAIE2, pero a un nivel más específico. El profesorado necesita entender claramente cómo las herramientas de IA encajan en su diseño instruccional. Ejemplos de integración de IA en educación se pueden encontrar en disciplinas como la ingeniería (Pereira et al., 2025) o la medicina (Hwang et al., 2024). Por ejemplo, en una clase de medicina donde el estudiantado está aprendiendo a diagnosticar y tiene a su disposición una herramienta de IA. Esta herramienta podría analizar datos de síntomas y sugerir posibles diagnósticos. Sin embargo, para alinearse con las prácticas didácticas, la IA no debería simplemente dar un diagnóstico final. En su lugar, debería funcionar como un asistente que guía a través de un proceso de razonamiento clínico, ayudando a identificar los síntomas más importantes, entender los posibles diagnósticos y considerar las decisiones de tratamiento. Esto permite que la IA complemente el diseño instruccional del profesorado y apoye el aprendizaje sin reemplazar el proceso de razonamiento del estudiantado.

➢El principio SAIE4 (precisión y minimización de errores) es crucial en la educación. Dado que las alucinaciones son inherentes al estado actual de las tecnologías de IA, es necesario un esfuerzo especial para minimizar las respuestas erróneas o engañosas. Esto se logra mejor en contextos de aplicación bien definidos y referenciando las fuentes usadas para generar las respuestas, lo que permite validar la información (Towhidul Islam Tonmoy et al., 2024).

➢El principio SAIE5 (interfaz comprensible y comportamiento adecuado) destaca la importancia de experimentar con interfaces que aclaren los usos y limitaciones de las herramientas de IA, evitando comportamientos que puedan inducir a error, como respuestas incorrectas con excesiva confianza.

➢Finalmente, los principios SAIE6 (supervisión humana) y SAIE7 (entrenamiento ético y transparencia) resaltan la necesidad de mantener siempre la supervisión humana en los procesos de IA, asegurar la responsabilidad ética y minimizar los sesgos en los modelos utilizados. Este enfoque de "IA Segura en la Educación" pone el énfasis en que la IA debe integrarse en los entornos educativos de forma que apoye y mejore la experiencia de enseñanza y aprendizaje, mientras se previene su mal uso y se atienden las preocupaciones éticas.

Estos principios han dado lugar a una discusión académica que ha culminado en la publicación del Safe AI Manifesto (Alier-Forment et al., 2024), un documento en continua evolución que se puede visitar y suscribir onli*ne* en https://manifesto.safeaieducation.org/.

3. LAMB. UNA PROPUESTA PARA INCORPORAR ASISTENTES DE IA EN EDUCACIÓN DE FORMA SEGURA

3.1. Asistentes de IA

Desde la disponibilidad generalizada de tecnologías de aprendizaje automático a mediados de la década de 2010, en particular con el desarrollo de Application Programming Interfaces (API) de alto nivel como TensorFlow (introducido en 2015 – https://www.tensorflow.org/) y PyTorch (introducido en 2016 – https://pytorch.org/), ha surgido una nueva categoría de software: el asistente. Un asistente es un sistema conversacional que utiliza interfaces de texto o voz y emplea tecnología de procesamiento de lenguaje natural (NLP – Natural Language Processing) para proporcionar acceso a un número limitado de características del sistema.
Cuando un chatbot basado en un LLM se combina con un conjunto de funciones de software tradicional se obtiene lo que se describe como un asistente de IA o asistente inteligente. Desde finales de 2022, las búsquedas de "asistente de IA" en los motores de búsqueda han aumentado casi diez veces, indicando la aparición de un nuevo tipo de asistente de IA implementado sobre modelos de LLM. Un ejemplo de asistente de IA es el motor de búsqueda perplexity.ai (https://perplexity.ai), que proporciona una respuesta al usuario basada en el contenido de las páginas obtenidas por la búsqueda y ofrece referencias dentro de su respuesta a los enlaces obtenidos. Utiliza la tecnología de LLM de la mejor manera posible: procesamiento de lenguaje natural, análisis de contenido, resumen y generación de contenido para proporcionar una respuesta completa con enlaces específicos y citas que permiten verificar la respuesta con fuentes autorizadas.

Un asistente de IA se diferencia de un chatbot basado en LLM en su propósito y comportamiento. Utiliza fuentes de información confiables, que puede citar y vincular con precisión, mientras mantiene las capacidades de procesamiento de lenguaje natural, razonamiento y seguimiento de instrucciones de un LLM. Es importante señalar que la calidad de un asistente no se basa únicamente en el volumen de información incorporada en el entrenamiento del LLM o en la fecha de actualización de su conocimiento. En cambio, depende de su capacidad para manejar contextos razonablemente grandes, recuperar información relevante y darle sentido, seguir instrucciones y estructurar salidas. Por lo tanto, el LLM ideal para construir un asistente de IA podría no ser necesariamente el modelo de mejor rendimiento en todas las áreas.

3.2. Tecnologías involucradas en la creación de un asistente de IA

Si bien los LLM desempeñan un papel crucial en el desarrollo de asistentes de IA (en adelante, "asistente"), no son la única tecnología emergente involucrada. Como se describe en la Tabla 1, hay una serie de tecnologías y disciplinas que contribuyen a crear un asistente integral. Estos componentes complementarios aumentan las capacidades del LLM, añadiendo nuevas características, previsibilidad y alcance.

Tabla 1 Lista de tecnologías involucradas en la creación de asistentes de IA.

Tecnología	Descripción
Generación aumentada por recuperación (Retrieval-Augmented Generation – RAG)	Combina la capacidad de los LLM para generar respuestas y obtener información de bases de datos o documentos externos, mejorando la precisión y relevancia de las respuestas. Los datos recuperados se insertan en la conversación con el LLM, generalmente en el llamado Contexto, para que el LLM pueda usarlos para generar una respuesta precisa (Lewis et al., 2020)
Búsqueda semántica en bases de datos de embeddings	Un embedding es una representación numérica de datos, de forma que se captura su significado y relaciones en un espacio de alta dimensión. Se utilizan embeddings para organizar y recuperar información de manera semántica, mejorando la capacidad del asistente para comprender y responder a consultas de forma precisa en diferentes modalidades. Una base de datos de embeddings puede realizar una búsqueda de similitud y recuperar objetos semánticamente relacionados con una consulta dada (Gao et al., 2024).
Contextos muy amplios	Los LLM modernos han comenzado a permitir contextos extensos, donde el contexto es la ventana de atención de un LLM en una conversación. Los LLM muestran la capacidad emergente de aprender nuevas habilidades a partir de la información y ejemplos proporcionados en una conversación
Intérpretes de código	Los LLM no están diseñados para realizar cálculos o tareas complejas. Sin embargo, cada vez son más hábiles en generar código que se puede pasar a un intérprete y luego utilizar la salida de la ejecución para completar su respuesta

Llamada a función	La llamada a función es una característica introducida por OpenAI en junio de 2023 (OpenAI, 2024). Permite al LLM responder con una invocación a la función de una API definida en el contexto. Esto facilita que el LLM interactúe con sistemas de información externos según los comandos del usuario
Ingeniería de prompts	La ingeniería de prompts (Sahoo et al., 2024) se refiere a la elaboración y optimización de prompts para guiar las respuestas del LLM, mejorando la calidad y relevancia del contenido generado
Marcos de evaluación y métricas	Uso de herramientas y técnicas para evaluar el rendimiento del asistente, incluyendo sistemas de monitoreo, métricas de precisión y conjuntos de datos de referencia, para garantizar su fiabilidad y efectividad
Ajuste fino (fine-tuning)	Se refiere al reentrenamiento del LLM subyacente en conjuntos de datos específicos para mejorar su rendimiento en dominios o tareas particulares (Christiano et al., 2023; Ouyang et al., 2022)

Fuente: elaboración propia

Más allá de las tecnologías listadas en la Tabla 1, los asistentes requieren habilidades de ingeniería de software y buenas prácticas, con un enfoque en el despliegue, escalabilidad y seguridad. Asegurar la seguridad de un asistente requiere una comprensión profunda de los principios de seguridad de la información y abordar preocupaciones únicas que surgen al incorporar un LLM en la pila tecnológica. Debe prestarse especial atención a mitigar riesgos como la inyección de prompts y el jailbreak de LLM (técnicas para intentar evadir las protecciones éticas y de seguridad), entre otras posibles amenazas (Liu & Hu, 2024; Yao et al., 2024).

3.3. Marco educativo para la creación de asistentes de aprendizaje basados en IA.

Se ha diseñado, desarrollado y evaluado un marco de software para el sector educativo que permita al profesorado crear asistentes de IA sin necesidad de programación, además de poder desplegarlos dentro del ecosistema tecnológico oficial de sus instituciones educativas. Este tipo de asistente de IA recibe el nombre de asistente de aprendizaje. El marco de software propuesto es el Learning Assistant Manager and Builder (LAMB) (Alier, Pereira, et al., 2024). El contexto académico implica que los temas de seguridad y ética son fundamentales. Por tanto, se debe examinar las implicaciones de la seguridad de la IA en el contexto de la educación para el diseño de LAMB y también los requisitos y restricciones específicos que el contexto educativo introduce. La exploración de los conceptos de IA

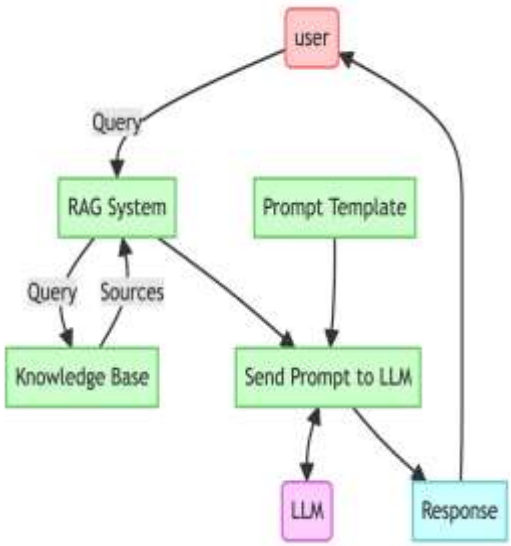
Segura en Educación procede de (Alier, García-Peñalvo, et al., 2024) y de las Aplicaciones de Aprendizaje Inteligente (Smart Learning Applications – SLA) (Alier, Casañ, et al., 2024).

Un asistente de aprendizaje debe ser interoperable con los Sistemas de Gestión de Aprendizaje (Learning Management Systems – LMS) y presentarse a estudiantes y profesorado como parte del ecosistema tecnológico institucional educativo. El cumplimiento de las políticas de privacidad de la institución educativa es un requisito legal en el sector educativo. También es necesaria la alineación del asistente de aprendizaje con la cultura de la institución educativa, utilizando solo fuentes de información autorizadas y herramientas para la generación de contenido, proporcionando citas adecuadas y transparencia.

3.4. Componentes de LAMB

La arquitectura de LAMB se basa en varios componentes clave que trabajan juntos para proporcionar respuestas precisas y contextualizadas a los usuarios. Cuando un usuario envía una pregunta, el sistema RAG recupera información relevante de la base de conocimiento, un repositorio de fuentes confiables que el sistema utiliza como contenido autorizado. A continuación, se utiliza la plantilla de prompt para estructurar la consulta y la información relevante en un formato que el LLM pueda procesar de manera efectiva. Este prompt combinado se envía al LLM, que genera una respuesta basada tanto en la pregunta del usuario como en los datos recuperados de la base de conocimiento. Finalmente, la respuesta se entrega al usuario, asegurando que el asistente proporcione respuestas precisas y bien fundamentadas, adaptadas al contenido educativo disponible (ver Figura 1).

Figura 1. Esquema de un asistente de aprendizaje simple.



Fuente: elaboración propia.

Ahora se va a examinar cómo funciona este asistente en un contexto educativo. El asistente recibe una "consulta" del usuario, por ejemplo:

"¿Cuáles son las causas de la Revolución Industrial?"

Esta consulta será procesada por el subsistema RAG del asistente, que usará la pregunta del usuario para seleccionar varios textos de su base de conocimiento. Se va a suponer que la base de conocimiento incluye materiales educativos aprobados, como artículos y libros de historia. El sistema obtiene un enlace a un artículo sobre la Revolución Industrial, además de un extracto que explica las principales causas, como los avances tecnológicos, la disponibilidad de recursos naturales y las transformaciones económicas de la época. Este texto y el enlace se consideran fuentes de información válidas. Las fuentes de información se combinan con la consulta del usuario, según una plantilla, para crear un mensaje para el LLM, por ejemplo:

"Responde a la pregunta: - ¿Cuáles son las causas de la Revolución Industrial? - considerando estas fuentes: Artículo de historia -enlace- Avances tecnológicos: desarrollo de máquinas de vapor Disponibilidad de recursos: acceso a carbón y hierro Transformaciones económicas: expansión del comercio y crecimiento de las ciudades -"

El LLM generará una respuesta como la siguiente, que se enviará al usuario:

"Las causas de la Revolución Industrial incluyen, entre otras, los avances tecnológicos como la máquina de vapor, la disponibilidad de recursos naturales como el carbón y el hierro, y las transformaciones económicas como la expansión del comercio y el crecimiento urbano. Estos factores facilitaron el desarrollo de las primeras industrias en Inglaterra, transformando profundamente la economía y la sociedad de la época."

Los asistentes de aprendizaje creados con LAMB se integran en el LMS Moodle utilizando el estándar de interoperabilidad de herramientas de aprendizaje (Learning Tools Interoperability – LTI) (IMS-GLC, 2014).

IMS LTI, desarrollado por el IMS Global Learning Consortium, permite que diferentes LMS se integren sin problemas con herramientas y contenidos educativos externos. Esto significa que herramientas como cuestionarios, tareas y otros recursos de aprendizaje de diversas plataformas pueden funcionar conjuntamente dentro de un único LMS, proporcionando una experiencia más fluida para los usuarios.

Por tanto, este protocolo permite que los asistentes se comporten como herramientas externas dentro de Moodle, manteniendo la autenticación y el contexto del usuario. Para configurar un asistente de aprendizaje en un curso de Moodle, el creador del asistente utiliza LAMB para generar una clave de asistente y un proveedor LTI, que luego se configura en Moodle como una actividad de herramienta externa. A través de esta integración, el estudiantado puede interactuar con el asistente directamente en el entorno de Moodle, donde se le garantiza que los datos se mantendrán privados y en conformidad con las políticas de la institución.

3.5. LAMB como IA segura en educación

Ahora se va a analizar este asistente de aprendizaje simple de acuerdo con los principios de IA Segura en Educación (SAIE).

El LLM se utiliza como una llamada de API a través del código del asistente. Esto significa que, a menos que el usuario decida incluir información personal en la consulta, se garantiza la confidencialidad del usuario. Esto satisface (SAIE1); garantiza la confidencialidad.

El sistema RAG utiliza una base de conocimiento que el profesor o la institución educativa proveen. Esto proporciona alineación con los estándares de calidad de la institución educativa, su visión sobre el tema y sus valores, cumpliendo con el SAIE2. La plantilla de prompt determinará el comportamiento del LLM, no la consulta del usuario. Esto satisface (SAIE3); está alineado con las prácticas didácticas y ayuda a cumplir con (SAIE5), presentando una interfaz y un comportamiento coherentes.

Según las indicaciones de la plantilla de prompt, "Responde esta pregunta <pregunta> según estas fuentes - <fuentes>", el LLM va a utilizar la información autorizada proporcionada por la base de conocimiento. El LLM basará sus respuestas en las fuentes proporcionadas, por lo que la precisión de la respuesta dependerá de la calidad de las fuentes recuperadas y no del entrenamiento del modelo ni de la fecha de actualización de su conocimiento. Esto satisface nuevamente (SAIE3), ya que está alineado con prácticas didácticas, y también cumple con el principio (SAIE4), que garantiza precisión y minimización de errores.

La institución educativa y el profesor diseñan y controlan el asistente al seleccionar y curar la base de conocimiento y al diseñar los prompts para definir el comportamiento del asistente. El asistente amplía y complementa las capacidades y funciones del profesorado y de la institución y no está diseñado para reemplazarlos, en total cumplimiento con el principio SAIE6.

Aunque el uso de un asistente no satisface directamente el principio SAIE7, ya que utiliza tecnología de backend basada en un LLM a través de API, sin tener control sobre el entrenamiento del modelo, el asistente en sí genera un conjunto de datos de interacciones —preguntas y respuestas— que luego pueden ser analizadas, verificadas y corregidas para crear un conjunto de datos de ajuste fino que la institución educativa puede usar para personalizar los LLM en el futuro. Es importante que este conjunto de datos futuro esté en manos de la institución educativa.

El análisis anterior sugiere que un asistente de IA adaptado a prácticas didácticas específicas puede satisfacer los principios SAIE1, 2, 3, 4, 5, 6, y, de manera indirecta, el 7. Sin embargo, para satisfacer completamente el SAIE2 (alineación con las estrategias educativas), el asistente de IA debe comportarse como una aplicación de aprendizaje inteligente. Esto se puede lograr utilizando el protocolo de interoperabilidad IMS LTI (IMS-GLC, 2014).

4. CASO PRÁCTICO DE APLICACIÓN DE LAMB

El curso de negocios es una asignatura básica del Grado en Ingeniería Informática de la EPSEVG (Escola Politècnica Superior de Vilanova i la Geltrú). Se imparte en el cuarto semestre, de febrero a junio, y tiene un valor de 6 créditos ECTS. En la edición del curso de este caso participaron 47 estudiantes.

En este curso se utilizó un asistente de aprendizaje para ayudar a los estudiantes a realizar una actividad de clase (Casañ et al., 2025). A continuación, se presenta el enunciado de la actividad que realizaron los estudiantes, así como la metodología utilizada y los resultados obtenidos de la utilización de este asistente.

El estudio de caso usado en el curso lleva por título: "Optimus, ¿el Transformer de Tesla?".

En octubre de 2022, Tesla presentó un robot humanoide llamado Optimus en el evento Tesla AI Day 2022, liderado por Elon Musk, quien cree que esta tecnología puede cambiar millones de vidas en el mundo.

Musk mostró un prototipo del robot, que utiliza el sistema de conducción autónoma de Tesla. Optimus caminó lentamente por el escenario, saludó al público y mostró algunos movimientos de baile. Musk afirmó que el robot podría hacer mucho más, pero no querían que se cayera. Agregó que Optimus podría ayudar a "millones" y transformar la civilización. También señaló que, aunque ahora se enfoca en trabajos en fábricas, en el futuro podría realizar tareas domésticas y hacer mandados. El precio estimado sería alrededor de $20,000 y estaría disponible en tres a cinco años.

El ejercicio consiste en analizar esta idea usando el método PESTLE (Rastogi & Trivedi, 2016; Zahari & Romli, 2019).

El caso de Optimus se desarrolló en el módulo de marketing, para contextualizar el análisis del entorno y el desarrollo de un análisis DAFO (Debilidades, Fortalezas, Amenazas y Oportunidades,). Se trabajó durante 2 horas en dos sesiones, con equipos de 6-8 miembros. Se les dio una semana entre las sesiones para seguir analizando el caso fuera del aula. Se siguieron las etapas siguientes:

- Paso 1. Introducción al caso y metodología (Sesión 1). El profesorado presentó el caso y explicó los fundamentos del análisis PESTLE, junto con los entregables esperados. Este paso duró entre 30 y 45 minutos.
- Paso 2. Colaboración en equipo (Sesión 1). El estudiantado se organizó en grupos pequeños (4-5 personas) para analizar el caso usando las seis dimensiones del método PESTLE. Debían elaborar un documento con los aspectos clave. Este paso tomó entre 45 y 60 minutos.
- Paso 3. Colaboración en equipo (Sesión 2). En la segunda sesión, se introdujo el análisis DAFO y se pidió que lo integraran con el análisis PESTLE. Cada elemento

del DAFO debía ser clasificado según su relevancia. Esta fase también tomó entre 30 y 60 minutos.

- Paso 4. Discusión grupal. Este paso es opcional y consiste en compartir las ideas con la clase. Un portavoz de cada grupo resume la evaluación de su equipo y se hace una discusión final moderada por el profesorado.

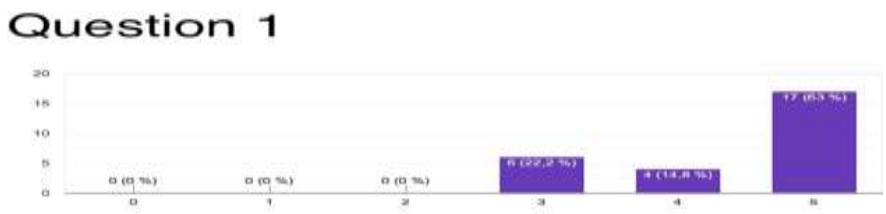
Aunque esta actividad se ha aplicado durante varios años académicos, esta es la primera vez que se utilizó un asistente de aprendizaje para que el estudiantado pueda hacer preguntas y obtener respuestas de expertos. El cambio consiste en que no solo buscan información en Internet sobre las dimensiones de PESTLE, sino que también tienen a su disposición un asistente que tiene una base de conocimiento de expertos sobre el tema al que pueden hacer preguntas. El asistente responde las preguntas y ofrece un enlace a las fuentes de datos. Además, el asistente también sugiere preguntas nuevas que pueden resultar de interés.

Al final de la actividad, se intentó determinar cómo percibía el estudiantado el estudio de caso, en términos de valor percibido. Los comentarios indican una recepción positiva de la herramienta LLM Mentor (basada en el framework LAMB), el asistente de aprendizaje utilizado. Por ello, se utilizó un breve cuestionario (ver Tabla 2). Los ítems de la encuesta fueron valorados por el estudiantado entre 0 y 5, siendo 0 la valoración más desfavorable y 5 la más favorable. Los resultados se pueden ver en las Figuras 2-5.

Tabla 2. Ítems de la encuesta realizada a los estudiantes.

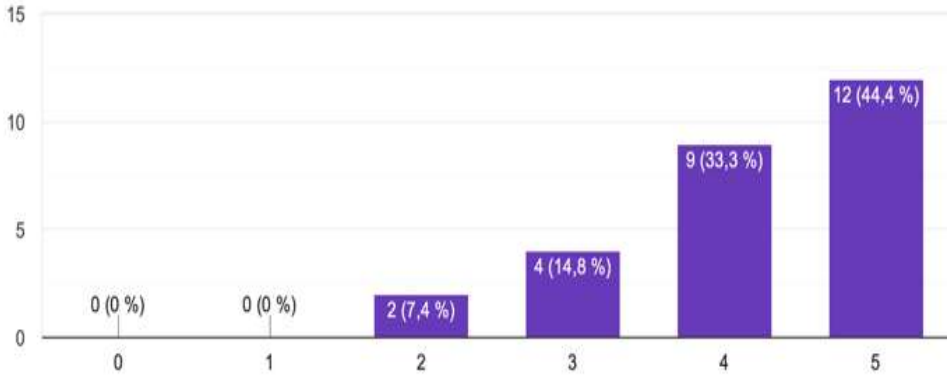
1. LLM Mentor me ha ayudado a encontrar información relevante para el caso más rápidamente que si hubiera tenido que hacerlo yo mismo usando internet.
2. Las preguntas adicionales sugeridas por el sistema son buenos puntos de partida para una búsqueda de información adicional sobre el caso.
3. Las respuestas a las preguntas sugeridas por el sistema proporcionan información útil.
4. Poder consultar fuentes de datos ha sido de ayuda.

Figura 2. Resultados ítem 1 "LLM Mentor me ha ayudado a encontrar información relevante para el caso más rápidamente que si hubiera tenido que hacerlo yo mismo usando internet".



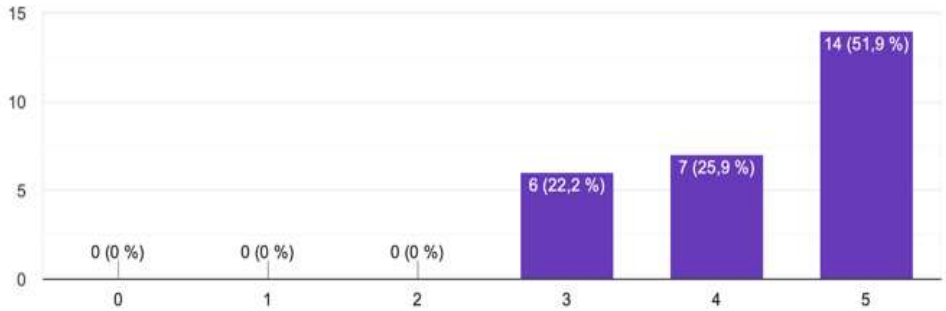
Fuente: elaboración propia.

Figura 3. Resultados ítem 2 "Las preguntas adicionales sugeridas por el sistema son buenos puntos de partida para una búsqueda de información adicional sobre el caso".



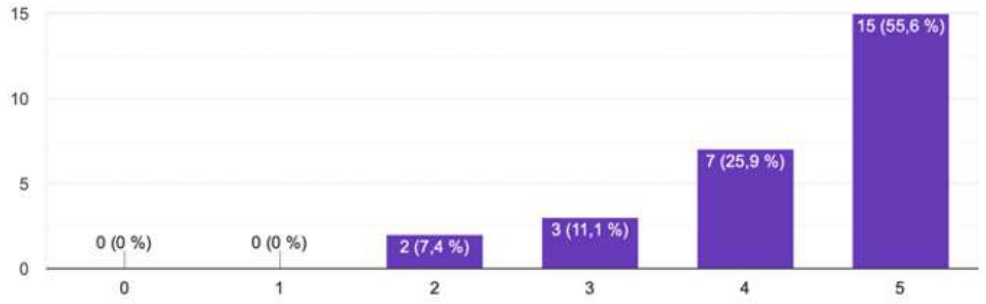
Fuente: elaboración propia.

Figura 4. Resultados ítem 3 "Las respuestas a las preguntas sugeridas por el sistema proporcionan información útil".



Fuente: elaboración propia

Figura 5. Resultados ítem 4 "Poder consultar fuentes de datos ha sido de ayuda".



Fuente: elaboración propia.

La media del ítem 1 (4,41) (ver Figura 2) indica que la mayoría del estudiantado consideró que la herramienta LLM Mentor fue útil para encontrar información de forma más rápida. La desviación estándar (0,84) muestra una ligera variabilidad en las respuestas, aunque en general las opiniones son bastante consistentes. En el ítem 2 la media (4,15) (ver Figura 3) sugiere que los estudiantes valoraron positivamente la calidad de las preguntas proporcionadas por el sistema. La desviación estándar (0,95) indica, de nuevo, una variabilidad moderada en las respuestas, con algunas opiniones que difieren de la mayoría. La media del ítem 3 (4,30) (ver Figura 4) sigue en la línea de los dos ítems anteriores. La desviación estándar (0,82) sugiere que las respuestas son bastante homogéneas de nuevo. Finalmente, en el último elemento, la media (4,30) (ver Figura 5) refleja que el estudiantado encontró valioso el acceso a las fuentes de datos que ofrece la herramienta. La desviación estándar (0,95) muestra que, aunque la mayoría valora positivamente este aspecto, hay algo de diversidad en las respuestas. En resumen, esta encuesta muestra que, en general, las preguntas relacionadas con la utilidad y calidad de la información de la herramienta recibieron respuestas favorables, aunque con algunas variaciones en las opiniones.

5. CONCLUSIONES

Integrar IAGen en educación, sea como estrategia de innovación o como reacción a la mera existencia de esta y su disponibilidad tanto para estudiantes y docentes, conlleva la responsabilidad de hacerlo de forma segura. En este capítulo se han delineado siete principios accionables para evaluar la estrategia de integración tecnológica de IAGen en entornos educativos, que se recogen en el AI Safe Manifesto (Alier-Forment et al., 2024), abierto a firma y evolución por parte de la comunidad.

Hace 20 años, cuando la adopción de las tecnologías web alcanzó una masa crítica que hizo imposible no tener una estrategia de adopción de la web en los procesos educativos, las universidades y centros educativos en España y en la mayoría de países de la Unión Europea decidieron adoptar soluciones de software libre, como Moodle o Sakai (Alier et al., 2021). Esta decisión aportó varios beneficios. Por una parte, esto confería soberanía tecnológica, una garantía que el conocimiento y talento necesario para operar y desarrollar las tecnologías sobre la que se iban a basar futuras estrategias educativas iba a permanecer en la institución educativa o en el tejido empresarial local. Y, por otro lado, la participación de una gran cantidad de docentes en las comunidades de aprendizaje online, encuentros y congresos facilitaba una cohesión entre los desarrolladores de la tecnología (comunidades open source) y los que iban a aplicar estos productos en la práctica docente. Como consecuencia, las tecnologías educativas básicas (especialmente los LMS) han estado relativamente libres de los gigantes tecnológicos que monopolizan la industria. Esta tendencia se rompe durante la pandemia de la COVID-19, pero ese es tema queda para un desarrollo futuro.

Los productos de IAGen "llave en mano" ofrecidos por actores como Microsoft (que controla OpenAI) ((Metz et al., 2024) o Google, no cumplen con los principios propuestos de seguridad en este capítulo, además de suponer un endeudamiento tecnológico tremendo en caso de ser integradas como piezas estrategias de la actividad por cualquier institución educativa. El proyecto LAMB es una muestra de herramienta educativa que permite integrar IAGen de forma segura, controlada y personalizada por el profesorado y las instituciones educativas; aprovechando las funcionalidades extraordinarias de la IAGen como una pieza intercambiable de un sistema en manos de la institución.

AGRADECIMIENTOS

Esta investigación está parcialmente financiada por el Ministerio de Ciencia e Innovación a través del proyecto AvisSA referencia (PID2020- 118345RB-I00), el Departament de Recerca i Universitats de la Generalitat de Catalunya a través de la ayuda para grupos de investigación 2021 SGR 01412, y la Universidad del País Vasco/Euskal Herriko Unibertsitatea a través del contrato GIU21/037 dentro del programa «Convocatoria para la Concesión de Ayudas a los Grupos de Investigación en la Universidad del País Vasco/Euskal Herriko Unibertsitatea (2021)».

REFERENCIAS

Alam, A., & Mohanty, A. (2022). Foundation for the Future of Higher Education or 'Misplaced Optimism'? Being Human in the Age of Artificial Intelligence. In M. Panda, S. Dehuri, M. R. Patra, P. K. Behera, G. A. Tsihrintzis, S.-B. Cho, & C. A. Coello Coello (Eds.), Innovations in Intelligent Computing and Communication. First International Conference, ICIICC 2022, Bhubaneswar, Odisha, India, December 16-17, 2022, Proceedings (pp. 17-29). Springer International Publishing. https://doi.org/10.1007/978-3-031-23233-6_2

Alier, M., Casañ Guerrero, M. J., Amo, D., Severance, C., & Fonseca, D. (2021). Privacy and E-Learning: A Pending Task. Sustainability, 13(16), *Article* 9206. https://doi.org/10.3390/su13169206

Alier, M., Casañ, M. J., & Amo, D. (2024). Smart Learning Applications: Leveraging LLMs for Contextualized and Ethical Educational Technology. In J. A. de Carvalho Gonçalves, J. L. Sousa de Magalhães Lima, J. P. Coelho, F. J. García-Peñalvo, & A. García-Holgado (Eds.), Proceedings TEEM 2023: Eleventh International Conference on Technological Ecosystems for Enhancing Multiculturality (Bragança, Portugal, October 25–27, 2023) (pp. 190-199). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-1814-6_18

Alier, M., García-Peñalvo, F. J., & Camba, J. D. (2024). Generative Artificial Intelligence in Education: From Deceptive to Disruptive. *International Journal of Interactive Multimedia and Artificial Intelligence,* 8(5), 5-14. https://doi.org/10.9781/ijimai.2024.02.011

Alier, M., Pereira, J., Garcia-Peñalvo, F. J., Casañ, M. J., & Cabré, J. (2024). LAMB: An Open-Source Software Framework to Create Artificial Intelligence Assistants Deployed and Integrated into Learning Management Systems. *Computer Standards & Interfaces, 92, Article* 103940. https://doi.org/10.1016/j.csi.2024.103940

Alier-Forment, M., García-Peñalvo, F. J., Casañ, M. J., Pereira, J. A., & Llorens-Largo, F. (2024). Safe AI in Education Manifesto. Version 0.4.0. https://manifesto.safeaieducation.orgAlier-Forment, M., & Llorens-Largo, F. (2023). EP-31 Las Alucinaciones de ChatGPT con Faraón Llorens. In Cabalga el Cometa. https://bit.ly/3ZCNBVT

Anthropic. (2024, March 4). Introducing the next generation of Claude. Anthropic. https://d66z.short.gy/c9wJor

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, Opportunities, and Threats. arXiv, Article arXiv:2304.09103v1. https://doi.org/10.48550/arXiv.2304.09103

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., & Amodei, D. (2020). Language Models are Few-Shot Learners. arXiv, Article arXiv:2005.14165v4 https://doi.org/10.48550/arXiv.2005.14165

Casañ, M. J., Llorens, A., Alier, M., & Pereira, J. (2025). Using an AI based learning assistant for a PESTLE case study learning activity. In Proceedings of the 12h International Con*ference on Technological Ecosystems for Enhancing Multiculturality 2024 - TEEM 2024* (23-25 October 2024, Alicante, Spain). Springer.

Chen, O., Paas, F., & Sweller, J. (2023). A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element *Interactivity. Educational Psychology Review,* 35(2), Article 63. https://doi.org/10.1007/s10648-023-09782-w

Choi, E. P. H., Lee, J. J., Ho, M. H., Kwok, J. Y. Y., & Lok, K. Y. W. (2023). Chatting or cheating? The impacts of ChatGPT and other artificial intelligence language models on nurse education. Nurse Education Today, 125, Article 105796. https://doi.org/10.1016/j.nedt.2023.105796

Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2023). Deep reinforcement learning from human preferences. arXiv, Article arXiv:1706.03741v4. https://doi.org/10.48550/arXiv.1706.03741

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International,* 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice*, 20(3). https://doi.org/10.53761/1.20.3.02

Duong, C. D., Ngo, T. V. N., Khuc, T. A., Tran, N. M., & Nguyen, T. P. T. (2024). Unraveling the dark side of ChatGPT: a moderated mediation model of technology anxiety and technostress. Information Technology & People, In Press. https://doi.org/10.1108/ITP-11-2023-1151

European Parliament. (2024). Artificial Intelligence Act. (P9_TA(2024)0138). European Parliament Retrieved from https://d66z.short.gy/2fRVtE.

European Parliament, & Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://bit.ly/2O2juE9

Fichten, C., Pickup, D., Asunsion, J., Jorgensen, M., Vo, C., Legault, A., & Libman, E. (2021). State of the research on artificial intelligence-based apps for post-secondary students with disabilities. *Exceptionality Education International,* 31(1), 62–76. https://doi.org/10.5206/eei.v31i1.14089

Fonseca-Escudero, D., García-Peñalvo, F. J., Llorens-Largo, F., & Molina-Carmona, R. (2023, 18-20 de octubre de 2023). *¡Qué viene la IA! ¿Estoy preparada/o? VII Congreso Internacional sobre Innovación, Aprendizaje y Cooperación*, CINAIC 2023, Universidad Politécnica de Madrid, Madrid, España. https://doi.org/10.5281/zenodo.10050857

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv, Article arXiv:2312.10997v5. https://doi.org/10.48550/arXiv.2312.10997García-Peñalvo, F. J. (2023). The perception of Artificial Intelligence in educational contexts after the launch of ChatGPT: Disruption or Panic? Education in the Knowledge Society, 24, Article e31279. https://doi.org/10.14201/eks.31279

García-Peñalvo, F. J. (2024). Generative Artificial Intelligence and Education: An Analysis from Multiple Perspectives. Education in the Knowledge Society, 25, Article e31942. https://doi.org/10.14201/eks.31942

García-Peñalvo, F. J., Alier, M., Pereira, J., & Casañ, M. J. (2024). Safe, Transparent, and Ethical Artificial Intelligence: Keys to Quality Sustainable Education (SDG4). IJERI – International Journal of Educational Research and Innovation, In Press.

García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED: revista iberoamericana de educación a distancia, 27(*1), 9–39. https://doi.org/10.5944/ried.27.1.37716

García-Peñalvo, F. J., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. International Journal of Interactive Multimedia and Artificial Intelligence, 8(4), 7-16. https://doi.org/10.9781/ijimai.2023.07.006

González-Geraldo, J. L., & Ortega-López, L. (2024). Can AI fool us? University Students' Lack of Ability to Detect ChatGPT. Education in the Knowledge Society, 25, Article e31760. https://doi.org/10.14201/eks.31760

Henley, A., Battle, R., & Gollapudi, T. (2024, March 6). AI prompt engineering is dead. IEEE Spectrum. https://d66z.short.gy/AN4am2

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. arXiv, Article arXiv:2311.05232v2. https://doi.org/10.48550/arXiv.2311.05232

Hwang, G.-J., Tang, K.-Y., & Tu, Y.-F. (2024). How artificial intelligence (AI) supports nursing education: profiling the roles, applications, and trends of AI in nursing education research (1993–2020). Interactive Learning Environments, 32(1), 373-392. https://doi.org/10.1080/10494820.2022.2086579

IMS-GLC. (2014). IMS Learning Tools Interoperability LTI Implementation Guide v2.0. https://bit.ly/488vnzN

Iskender, A. (2023). Holy or Unholy? Interview with Open AI's ChatGPT. European Journal of Tourism Research, 34, Article 3414. https://doi.org/10.54055/ejtr.v34i.3169

Johinke, R., Cummings, R., & Di Lauro, F. (2023). Reclaiming the technology of higher education for teaching digital writing in a post—pandemic world. Journal of University Teaching and Learning Practice, 20(2), Article 01. https://doi.org/10.53761/1.20.02.01

Jovanović, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. Computer, 55(10), 107-112. https://doi.org/10.1109/MC.2022.3192720

Kamath, U., Keenan, K., Somers, G., & Sorenson, S. (2024). LLM Challenges and Solutions. In U. Kamath, K. Keenan, G. Somers, & S. Sorenson (Eds.), Large Language Models: A Deep Dive: Bridging Theory and Practice (pp. 219-274). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65647-7_6

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler†, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), Advances in Neural Information Processing Systems (pp. 9459-9474).

Liu, F. W., & Hu, C. (2024). Exploring Vulnerabilities and Protections in Large Language Models: A Survey. arXiv, Article arXiv:2406.00240v1. https://doi.org/10.48550/arXiv.2406.00240

Meta. (2024, April 18). Introducing Meta Llama 3: The most capable openly available LLM to date. Meta. https://d66z.short.gy/95zf7bMetz, C., Isaac, M., & Griffith, E. (2024, October 21). Microsoft and OpenAI's Close Partnership Shows Signs of Fraying. New York Times. https://d66z.short.gy/8pIG6E

Mistral AI team. (2024a, April 17). Cheaper, Better, Faster, Stronger. Continuing to push the frontier of AI and making it accessible to all. Mistral AI_. https://d66z.short.gy/cgRVBp

Mistral AI team. (2024b, November 18). Pixtral Large. Pixtral grows up. Mistral AI_. https://d66z.short.gy/9yv6oM

Morales-Chan, M. (2023). ChatGPT en la Investigación: Creando Prompts Efectivos. Universidad Galileo. https://bit.ly/ChatGPTInvestigacion

OpenAI. (2024). OpenAI Platform. https://d66z.short.gy/BiocOD

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. arXiv, Article arXiv:2203.02155v1. https://doi.org/10.48550/arXiv.2203.02155

Pereira, J., López-Gil, J. M., & Alier, M. (2025). The AI-Powered Classroom: LLMs as Teacher Assistants for Enhanced Software Engineering Learning Experiences. In Proceedings of the 12h International Conference on Technological Ecosystems for Enhancing Multiculturality 2024 - TEEM 2024 (23-25 October 2024, Alicante, Spain). Springer.

Perković, G., Drobnjak, A., & Botički, I. (2024). Hallucinations in LLMs: Understanding and Addressing Challenges. In 2024 47th MIPRO ICT and Electronics Convention (MIPRO) (Opatija, Croatia, 20-24 May 2024) (pp. 2084-2088). IEEE. https://doi.org/10.1109/MIPRO60963.2024.10569238

Pichai, S., & Hassabis, D. (2024). Our next-generation model: Gemini 1.5. AI. https://d66z.short.gy/cT19l1

Rastogi, N., & Trivedi, M. K. (2016). PESTLE technique. A tool to identify external risks in construction projects. International. Research Journal of Engineering and Technology, 3(1).

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. arXiv, Article arXiv:2402.07927v1. https://doi.org/10.48550/arXiv.2402.07927

Shetye, S. (2024). An evaluation of Khanmigo, a generative AI tool, as a computer-assisted language learning app. Studies in Applied Linguistics and TESOL, 24(1), 38-53. https://doi.org/10.52214/salt.v24i1.12869

Towhidul Islam Tonmoy, S. M., Mehedi Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models. arXiv, Article arXiv:2401.01313v3. https://doi.org/10.48550/arXiv.2401.01313

Vartiainen, H., & Tedre, M. (2023). Using artificial intelligence in craft education: crafting with text-to-image generative models. Digital Creativity, 34(1), 1-21. https://doi.org/10.1080/14626268.2023.2174557

Willison, S. (2023, February 21). In defense of prompt engineering. Simon Willison's Weblog. https://d66z.short.gy/APdhKn

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., & Hu, X. (2024). Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. ACM Transactions on Knowledge Discovery from Data, 18(6), Article 160. https://doi.org/10.1145/3649506Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly. High-Confidence Computing, 4(2), Article 100211. https://doi.org/10.1016/j.hcc.2024.100211

Zahari, A. R., & Romli, F. I. (2019). Analysis of suborbital flight operation using PESTLE. Journal of Atmospheric and Solar-Terrestrial Physics, 192. https://doi.org/10.1016/j.jastp.2018.08.006